

8-16-00

A

jc542 U.S. PTO
08/15/00UTILITY PATENT APPLICATION TRANSMITTAL
(Small Entity)

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Docket No.
FKC-100USTotal Pages in this Submission
47

TO THE ASSISTANT COMMISSIONER FOR PATENTS

Box Patent Application
Washington, D.C. 20231

Transmitted herewith for filing under 35 U.S.C. 111(a) and 37 C.F.R. 1.53(b) is a new utility patent application for an invention entitled:

ELECTRONIC MESSAGE PROCESSING

and invented by:

Patrick McErlean

jc542 U.S. PTO
09/15/00If a **CONTINUATION APPLICATION**, check appropriate box and supply the requisite information: Continuation Divisional Continuation-in-part (CIP) of prior application No.: _____

Which is a:

 Continuation Divisional Continuation-in-part (CIP) of prior application No.: _____

Which is a:

 Continuation Divisional Continuation-in-part (CIP) of prior application No.: _____

Enclosed are:

Application Elements

1. Filing fee as calculated and transmitted as described below
2. Specification having 32 pages and including the following:
 - a. Descriptive Title of the Invention
 - b. Cross References to Related Applications (*if applicable*)
 - c. Statement Regarding Federally-sponsored Research/Development (*if applicable*)
 - d. Reference to Microfiche Appendix (*if applicable*)
 - e. Background of the Invention
 - f. Brief Summary of the Invention
 - g. Brief Description of the Drawings (*if drawings filed*)
 - h. Detailed Description
 - i. Claim(s) as Classified Below
 - j. Abstract of the Disclosure

UTILITY PATENT APPLICATION TRANSMITTAL
(Small Entity)

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Docket No.
FKC-100US

Total Pages in this Submission
47

Application Elements (Continued)

3. Drawing(s) *(when necessary as prescribed by 35 USC 113)*
a. Formal b. Informal Number of Sheets 5
4. Oath or Declaration
a. Newly executed *(original or copy)* Unexecuted
b. Copy from a prior application (37 CFR 1.63(d)) *(for continuation/divisional application only)*
c. With Power of Attorney Without Power of Attorney
d. DELETION OF INVENTOR(S)
Signed statement attached deleting inventor(s) named in the prior application,
see 37 C.F.R. 1.63(d)(2) and 1.33(b).
5. Incorporation By Reference *(usable if Box 4b is checked)*
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
6. Computer Program in Microfiche
7. Genetic Sequence Submission *(if applicable, all must be included)*
 - a. Paper Copy
 - b. Computer Readable Copy
 - c. Statement Verifying Identical Paper and Computer Readable Copy

Accompanying Application Parts

8. Assignment Papers *(cover sheet & documents)*
9. 37 CFR 3.73(b) Statement *(when there is an assignee)*
10. English Translation Document *(if applicable)*
11. Information Disclosure Statement/PTO-1449 Copies of IDS Citations
12. Preliminary Amendment
13. Acknowledgment postcard
14. Certificate of Mailing
 First Class Express Mail *(Specify Label No.):* EL 711 312 842 US

**UTILITY PATENT APPLICATION TRANSMITTAL
(Small Entity)**

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Docket No.
FKC-100US

Total Pages in this Submission
47

Accompanying Application Parts (Continued)

15. Certified Copy of Priority Document(s) *(if foreign priority is claimed)*

16. Small Entity Statement(s) - Specify Number of Statements Submitted: _____

17. Additional Enclosures *(please identify below):*

--

Fee Calculation and Transmittal

CLAIMS AS FILED

For	#Filed	#Allowed	#Extra	Rate	Fee
Total Claims	13	- 20 =	0	x \$9.00	\$0.00
Indep. Claims	3	- 3 =	0	x \$39.00	\$0.00
Multiple Dependent Claims (check if applicable)					\$0.00
				BASIC FEE	\$345.00
OTHER FEE (specify purpose)				Recordation of Assignment	\$40.00
				TOTAL FILING FEE	\$385.00

2 checks in amount of \$345 & \$40 to cover the filing fee is enclosed.
 The Commissioner is hereby authorized to charge and credit Deposit Account No. 18-0350 as described below. A duplicate copy of this sheet is enclosed.
 Charge the amount of _____ as filing fee.
 Credit any overpayment.
 Charge any additional filing fees required under 37 C.F.R. 1.16 and 1.17.
 Charge the issue fee set in 37 C.F.R. 1.18 at the mailing of the Notice of Allowance, pursuant to 37 C.F.R. 1.311(b).

Dated: August 15, 2000


Signature

Kenneth N. Nigon, Reg. No. 31,549
RATNER & PRESTIA
Suite 301
One Westlakes, Berwyn
P.O. Box 980
Valley Forge, PA 19482-0980
(610) 407-0700

cc:

CERTIFICATE OF MAILING BY "EXPRESS MAIL" (37 CFR 1.10)Applicant(s): **Patrick McErlean**

Docket No.

FKC-100USSerial No.
To Be AssignedFiling Date
Herewith

Examiner

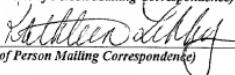
Group Art Unit

Invention: **ELECTRONIC MESSAGE PROCESSING**

I hereby certify that the following correspondence:

Trans.(3 pgs.); Specification(32 pgs.); Formal Drawings(5 sheets); Unexecuted Decl/POA(2 pgs.); Assignment w/cover sheet(4 pgs.); 2 checks; Postcard*(Identify type of correspondence)*

is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 in an envelope addressed to: The Assistant Commissioner for Patents, Washington, D.C. 20231 on

August 15, 2000*(Date)***Kathleen Libby***(Typed or Printed Name of Person Mailing Correspondence)**(Signature of Person Mailing Correspondence)***EL 711 312 842 US***("Express Mail" Mailing Label Number)*

Note: Each paper must have its own certificate of mailing.

Electronic Message Processing

5 FIELD OF THE INVENTION

The present invention relates to electronic message processing. The invention relates particularly, but not exclusively, to the categorization of electronic messages.

10

BACKGROUND TO THE INVENTION

Many organizations are now equipped to receive messages, including orders and enquiries, by electronic means.

15 Typically, such electronic messages take the form of text-based messages, for example e-mails, delivered by a global computer network, for example the Internet, or by a telecommunications network, for example a mobile telephone network. Each message must be processed and dealt with 20 appropriately. In many cases, the volume of received electronic messages is relatively high and it is considered inefficient to process each message manually.

For this reason, it is known for electronic message

25 processing systems, typically in the form of a computer system, to employ a text analyzer, such as IBM's Mail Analyzer, to analyze the content of electronic messages in order to classify, or categorize, each message according to its content. Once a message has been categorized, the 30 processing system sends it on to a human operator who has the skills necessary to deal with messages falling within the relevant category(s). Alternatively, the computer

system itself may be able to deal with messages falling within certain category(s).

Typically, a text analyzer examines the text of each 5 message in turn in search of one or more alphanumeric text string, for example a word or sequence of words, which may be used to identify the purpose or nature of the message under examination. It is known for the text analyzer to operate in association with a rule engine to apply a set 10 of rules to the message in order to determine how to categorize the current message.

By way of simplistic example, consider a banking organization which has a message processing system 15 arranged to receive electronic messages in three different categories, namely: balance enquiry; request for funds transfer; and interest rate enquiry. In order to categorize each received message, a text analyzer in association with a rule engine applies a set of four rules 20 to each message in turn. The first rule stipulates that if the text string "balance" appears in the message, then the message falls in the balance enquiry category. The second rule stipulates that if the text strings "funds" and "transfer" appear in the message, then the message 25 falls in the funds transfer category. The third rule stipulates that if the message contains the text string "interest rate", then the message should be categorized as an interest rate enquiry. The fourth rule stipulates that if none of the previous rules are satisfied, then the 30 message is deemed unclassified. Clearly, a message may fall within more than one category.

The message processing system may be arranged to distribute all balance enquiries, fund transfers and unclassified messages to an appropriate human operator, while being arranged to send out interest rate information 5 itself.

Such systems are suited to processing relatively small volumes of messages falling over a relatively small number of categories but exhibit serious shortcomings when 10 dealing with large volumes of messages and a large number of categories.

It is increasingly common for organizations to receive up to hundreds of thousands of electronic messages each day, 15 each message requiring classification into one or more of typically hundreds of different categories. To perform message categorization, a text analyzer would typically apply a set of several hundred rules to each message, the final classification of each message being derived from a 20 combination of the results of the application of all of the rules to that message. Conventionally, a flat rule structure is employed meaning that each rule is given equal weight and is applied in sequence to each message, one message at a time. This requires a large amount of 25 computer processing power which can lead to unacceptable delays in dealing with incoming messages.

Further, for a complex taxonomy, the precision (i.e. the degree to which messages are categorized correctly) and 30 recall (i.e. the degree to which a complete set of message categorizations are produced) are difficult to maintain at a high level.

It is also problematic to administer a single set of rules over a complex organization having a number of different divisions or sub-organizations since one or more categories may need to be defined or characterized

5 differently to account for differences in culture, regulation, market segmentation, brand specificity, or the like. Similar problems arise where more than one organization shares a single message processing system (and therefore a single set of rules) through, for

10 example, an Application Service Provider (ASP). There is a potential for conflict over the rules in that the rules for one organization, or sub-organization, may affect the application of the rules of another organization, or sub-organization, and so lead to inappropriate

15 categorizations.

This problem is compounded when the different organizations, or sub-organizations, need to be able to receive messages in different languages. A particular

20 problem that multiple languages causes concerns the performance of lexical analysis (sometimes known as word-stemming) on the message under examination. Lexical analysis is desirable since it enables the text analyser to recognize different forms of words, such as plurals and

25 participles, and so helps messages to be categorized correctly. In general, Lexical analysis is typically performed using a dictionary but a conventional text analyzer can only operate with one dictionary at a time. If, for example, a text analyzer is initialized with an

30 English dictionary, then any messages received in any other language cannot benefit from lexical analysis.

SUMMARY OF THE INVENTION

Accordingly, the present invention provides an electronic message processing system arranged to receive electronic messages, the system comprising: means for storing a plurality of rules; a text analyzer; a rule engine associated with text analyzer and with the rule storage means, the text analyzer and the rule engine being co-operable to apply one or more rules to the content of a received electronic message and to generate one or more results based on the application of the or each rule; a classification module co-operable with the text analyzer and the rule engine and arranged to classify the electronic message into one or more message categories based on said one or more results, wherein the rules are arranged into a plurality of rule sets, the classification module being arranged to cause the text analyzer in association with the rule engine to apply one or more rule sets to the message content in accordance with a hierarchical structure whereby the or each result generated by application of one rule set to the message content determines the next rule set, or rule sets, to be applied.

Arranging rules into multiple rule sets is advantageous in that each rule set can be maintained and administered independently i.e. without interfering with other rule sets. Thus, by allocating a respective rule set, or sets, to respective organizations or sub-organizations, the likelihood of conflict amongst organizations/sub-organizations is greatly reduced. Further, as a result of using the hierarchical structure, only those rule sets

that are relevant to the message at hand are applied by the rule set. This greatly reduces the processing power required to classify a message.

- 5 Preferably, the text analyzer includes the rule engine. More preferably, the classification module is arranged to instantiate a respective instance of the text analyzer for each rule set, each text analyzer instance being arranged to apply its respective rule set to the message content.
- 10 Further preferably, each text analyzer instance is associated with a respective lexical analysis tool, such as a dictionary.

This allows each message to be analyzed using a dictionary of an appropriate language and so the benefits of lexical analysis are not lost on any message.

- 15
- 20 Preferably, the rule storage means comprises a plurality of rule files, each rule file containing a respective rule set. This facilitates independent administration and maintenance of each rule set.

The invention also provides a classification module for use in an electronic message processing system, as claimed in claim 8.

The invention further provides a method of classifying an electronic message as claimed in claim 9.

- 25
- 30 The invention also provides an electronic mail (e-mail) processing system comprising the electronic message processing system of the invention. The invention further

provides an SMS message processing system comprising the electronic message processing system of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

5

An embodiment of the invention is now described by way of example and with reference to the accompanying drawings in which:

10 **Figure 1** is a schematic view of a communications network including a electronic message processing system;

Figure 2 is a more detailed schematic view of the electronic message processing system of Figure 1;

15

Figure 3 is a schematic view of a hierarchical structure for use, in accordance with the present invention, in the classification of electronic messages;

20

Figure 4 is a schematic view of a classification module and multiple instances of a text analyzing module arranged in accordance with the present invention; and

25

Figure 5 is a flow diagram illustrating the operation of the classification module of Figure 4.

DETAILED DESCRIPTION OF THE DRAWINGS

With reference to Figure 1 of the drawings, there is shown, generally indicated at 10, a communications network for the transmission of electronic messages 12. The network 10 includes a global computer network 14, such as the Internet, and a mobile telephone network 16 but, in

general, may include any computer network and/or telecommunications network, including WAN, LAN or MAN networks. The electronic messages 12 typically take the form of, for example, e-mails, Web-page forms, SMS (Short Message Service) text messages as commonly used by mobile telephones or the like. It will be understood that the present invention is not limited for use with any particular such type or types of electronic messages.

10 The electronic messages 12 are transmitted across the communications network 10 in conventional manner using, for example, a mail server 20, Web server 22, servlet 24 and/or WAP (Wireless Application Protocol) gateway 26, as appropriate, and using conventional transmission protocols such as, WAP, HTTP (Hyper Text Transfer Protocol), SMTP (Simple Mail Transfer Protocol), POP (Post Office Protocol), or other suitable protocol. It will be understood that the invention is not limited to use with any particular transmission means or protocols.

15

20 The electronic messages 12 are received by a message processing system 18 which may reside, for example, on an organization's computer system (not shown) or an ASP's computer system (not shown). The message processing

25 system 18 includes means for receiving electronic messages in conventional manner which, in the example of Figure 1, includes an e-mail receiving module 28, an SMS receiving module 30 and a database 32 for storing received messages 12.

30 In the example illustrated in Figure 1, the e-mail receiver 28 is arranged to receive e-mails in two ways. The first is directly from an ISP (Internet Service

Provider) using SMTP, the second is via the mail server 20 using POP3. The e-mail receiver 28 is further arranged to receive Web-page form messages through the web server 22. Web-page form messages may be considered as structured e-
5 mails which call the servlet 24 in the web server 22 which in turn converts the Web-page from content into an e-mail which is subsequently sent to the e-mail receiver 28 using SMTP.

10 Web-page forms from the mobile network 16 are transmitted using a WAP protocol stack to the WAP gateway 26. The WAP gateway 26 converts the Web-page form into an HTTP request, which is then processed by the web server 22 and servlet 24 as described above. SMS messaging from the
15 mobile network 16 is conveniently transmitted using WAP to the SMS receiver 30.

Thus, the e-mail receiver 28 and the SMS receiver 30 are arranged in conventional manner to support the appropriate
20 transmission protocols which, in the example of Figure 1, are SMTP and POP3 for the e-mail receiver 28, and WAP for the SMS receiver.

The communications network 10 of Figure 1 is given by way
25 of example only to provide a typical context for the invention and is not limiting to the invention.

Referring now to Figure 2, the message processing system
18 is shown in more detail. Electronic messages 12
30 received by the e-mail receiver 28 and SMS receiver 30 are stored in database 32. A classification module 34, or classifier, retrieves each message 12 from the database 32 in turn. The primary function of the classifier 34 is to

5 classify, or categorize, each message 12 into one or more categories depending on the content of the message 12. To this end, the classifier 34 is associated with, or includes, a text analyzer 36, for example the IBM Mail Analyzer and is further associated with, or includes, a rule engine. This classification process is described in more detail with reference to Figures 3, 4 and 5.

10 Once classified, each message is returned to the database 32 by the classifier 34 together with the determined classification information and, preferably, with information as to where the message should be routed in view of its classification. For example, messages 12 falling within certain categories may require to be dealt 15 with by a human operator (not shown) with particular skills, while messages 12 falling within other categories may be dealt with automatically by a virtual agent 38.

20 A distributor module 40, or distributor, retrieves the categorized messages and sends them either to the virtual agent 38 or a mail client 42, according to the respective categorization and/or routing information. The virtual agent 38 includes one or more software routines which automatically generate an appropriate response to a 25 message 12 according to its categorization information. The mail client 42 provides the message 12 to a human operator, or agent, who then prepares an appropriate response. A message 12 may, depending on the number and nature of categories into which it falls, be passed by the 30 distributor 40 amongst a plurality of human and/or virtual agents in order that a complete response may be prepared. Once a final response is prepared, the distributor 40

marks the response as complete and stores it in the database 32.

5 A composing module 44, or composer, retrieves responses from the database 32 that are marked as complete and composes a respective response message (not shown) for reply to the originator (not shown) of the respective received message 12. The composer 44 typically fulfils any attachment requirements as well as adding the 10 appropriate Internet, or other, protocol information. Each composed response message is then returned to the database 32.

15 An e-mail dispatch module 46, or dispatcher, retrieves from the database 32 the composed response messages which are destined to be transmitted by e-mail and conveniently dispatches them to an ISP (not shown) using SMTP. An SMS dispatch module 48, or dispatcher, retrieves from the database 32 the composed response messages which are 20 destined to be transmitted as SMS messages and dispatches them to the mobile network 16 using SMS protocol.

Conveniently, the modules of Figure 2 take the form of 25 CORBA (Common Object Request Broker Architecture) modules, or servers written, for example, in the Java programming language. Each module may reside on a single computer or may be distributed across a computer system, as is convenient. The e-mail receiver 28, SMS receiver 30, virtual agent 38, distributor 40, mail client 42, composer 30 44, e-mail dispatcher 46 and SMS dispatcher 48 are conveniently conventional in configuration and operation and are not limiting to the present invention.

The present invention is concerned primarily with the operation of the classifier 34 in association with the text analyzer 36 and rule engine. Some electronic messages 12 are structured i.e. they are composed in accordance with a specific known format, and may be inherently classified into one or more categories in cases where their purpose is dictated by the structure. For example, a particular Web-page form (not shown) is typically provided on the Web site of an organization in order for their customers to make a specific enquiry. In the example of a simple banking organization, one Web-page form may be used to make a balance enquiry while a different Web-page form may be used to make a request for a funds transfer. When the banking organization receives an electronic message 12 emanating from such a Web-page form, the nature or purpose of the message 12 is inherently known by virtue of the Web-page form used. In other cases, however, received electronic messages 12 are unstructured and their nature or purpose is not known inherently. The function of the classifier 34, in association with the text analyser 36, is to determine the purpose or nature of an electronic message 12 by examining its content and subsequently classifying the message 12 into one or more categories which reflect the determined purpose or nature.

The text analyzer 36 is preferably of the type which includes a rule engine (not shown). The IMB Mail Analyzer is an example of such a text analyser 36. The text analyzer 36 is therefore associated with a set of rules which it uses, in conjunction with the rule engine, to evaluate the content of an electronic message 12. In an alternative embodiment, the text analyzer and the rule

engine may be separate modules arranged for co-operation with one another. Typically, a rule comprises a conditional statement whereby an assertion is made if a condition is met. The condition is typically the presence 5 in the message content of a particular alphanumeric text string or character string. The assertion is typically that the message 12 belongs, or may belong, to a particular category, or categories. Thus, to evaluate a rule, the text analyzer examines the message content, 10 using conventional text analysis methods, searching for the presence of the particular string, or strings, of text upon which said rule is conditional. In the simplistic banking example, one rule may stipulate that if the text string "balance" appears in the message 12, then the 15 message 12 falls in the balance enquiry category. A second rule may stipulate that if the text strings "funds" and "transfer" appear in the message 12, then the message 12 falls in the funds transfer category, and so on. In practice, there may be hundreds of rules in a set and more 20 than one rule may be required to be met before the text analyser 36 determines that a message 12 falls within a particular category.

The rules themselves may be determined manually but are 25 conveniently generated automatically using a training tool (not shown), such as the IBM Mail Analyzer Training Tool. The training tool is provided by an operator (not shown) with a batch of sample messages which are known to fall within a particular category. The training tool 30 examines each sample message in the batch looking for characterizing alphanumeric text strings such as a word or group of words. The training tool uses these characterizing text strings to generate one or more rules

which may subsequently be used by the text analyzer 36 when evaluating each real message 12. The process is repeated for each category used by the, or each, organization, or sub-organization, which uses the same 5 message processing system 18. The number of categories, and therefore the number of rules, depends on the organization(s) but may, for example, number several hundred.

10 Conventionally, the classifier 34 is associated with only one instance of a text analyzer 36 and the single text analyzer 36 instance is associated with only one rule set and, optionally, one dictionary (to allow the text analyzer 36 to perform a lexical analysis, or 'word- 15 stemming'). Hence, all of the rules generated for all of the categories are conventionally stored in a single rules file (not shown in Figure 2) which is associated with the text analyzer 36 upon its initialization. In use, each message 12 is evaluated against each rule in sequence 20 (flat rule structure). The problems with operating a single rules file and a flat rule structure where there are a relatively large number of categories and rules and/or more than one organization, or sub-organization, to cater for, are outlined above.

25 Referring now to Figure 3, in accordance with the present invention a hierarchy, or hierarchical structure 310, is employed in the classification, or categorization, of electronic messages 12. The hierarchical structure 310, 30 which may also be referred to as a decision tree, comprises a plurality of nodes arranged in a hierarchy. The nodes include a root node 311, one or more branch nodes 312 and one or more leaf nodes 316. The root node

311 and the or each branch node 312 is associated with a respective set of one or more rules 314. A message 12 to be classified is first subject to the rule set 314 associated with the root node 311 which, in Figure 3, is
5 NODE 1. The result of the NODE 1 rule set evaluation determines under which branch node 312 the message 12 is subsequently evaluated which, in the example hierarchy of Figure 3, is either NODE 2 or NODE 3. Evaluation of the rule set 314 associated with NODE 2 determines whether the
10 message 12 should subsequently be evaluated under a further branch node 312 (NODE 4) and/or classified as category A at a leaf node 316. Similarly, evaluation of the rule set 314 associated with NODE 3 determines whether the message 12 should subsequently be evaluated under NODE
15 5 and/or classified as category B. A leaf node 316 terminates a branch of the hierarchy and corresponds to a respective classification category. In general, the rule set 314 of a given root or branch node 311, 312 determines, when applied to a message 12, under which
20 further branch node 312, and therefore which further rule set, the message 12 is subsequently to be evaluated and/or at which leaf node 316 the message 12 is classified. Each message is effectively routed through the hierarchy until it has been classified into one or more categories which,
25 in the example of Figure 3, may be one or more of categories A to F.

The hierarchy in Figure 3 is given by way of example only. The actual shape and size of the hierarchy (i.e. the
30 number and type of nodes and their inter-dependencies) will depend on the number and nature of the possible categories and therefore on the needs of the organization(s) who use the message processing system. In

general, one or more branch nodes 312 may hang from, or depend on, a root node 311 or preceding branch node 312, and one or more leaf nodes 316 may hang from, or depend on, a root node 311 or preceding branch node 312. A 5 message 12 falling within more than one category may be routed through the hierarchy along more than one branch to arrive at the relevant leaf node(s) 316.

The electronic message processing system of the invention 10 includes a classification module which stores, or has access to, the hierarchical structure to be implemented. The classification module is arranged to associate each of the branch nodes and the root node with a respective rule set and to associate each leaf node with one or more 15 respective message category, and to cause the text analyzer, to apply (by means of its rule engine), to the content of a message, firstly the rule set associated with said root node and then to determine, depending on the or each result generated by the application of said rule set, 20 the next node in the path or route through the hierarchy i.e. whether the next node is a branch node and, if so, which branch node, or a leaf node. If a branch node is encountered, then the classification module causes the text analyzer to apply the rule set associated with that 25 branch node. The process repeats until the hierarchical structure determines that the next node is a leaf node, in which case classification of the message is complete i.e. an appropriate category for the message has been found. As is described in more detail with reference to Figure 5, 30 the classification module may navigate along more than one route through the hierarchical structure 310. In Figure 3, for example application of the rule set associated with branch NODE 2 may generate a result set of, say, two

results. One result may indicate that the message falls into category A. The other result may indicate that the message content should next be evaluated using the rule set associated with NODE 4. The classification module

5 continues to apply rule sets to the message content until each path through the hierarchy reaches a leaf node 316 - this indicates that classification, or categorization, of the message is complete. Thus, in general, the classification module is arranged to cause the text

10 analyzer to apply a one or more rule sets to the message content in accordance with a hierarchical structure whereby the or each result generated by the rule engine upon application of one rule set to the message content determines the next rule set, or sets, to be applied.

15 Effectively, the rules are arranged into rule sets, each set having one or more rules, and the rule sets are arranged according to a hierarchical structure which is implemented, by the classification module, as a decision tree such that the next node to be implemented depends on

20 the outcome of the application of the rule set of the previous node. The rules may be stored in the system by any convenient means, for example in a computer file, in memory or in a database. Preferably, each rule set is stored in a respective computer file since this

25 facilitates the administration and maintenance of each rule set independently of the others. Similarly, the data describing the hierarchical structure can be stored by any conventional means, such as a computer file.

30 An example implementation of the hierarchical structure 310 of Figure 3 is now described with reference to Figures 4 and 5. In Figure 4, there is shown a classification module 434, or classifier, in association with a text

analyzer 436 (which includes a rule engine (not shown))
for use in an electronic message processing system
according to the present invention. The classifier 434 is
co-operable with a plurality of instances 436' of the text
5 analyzer 436, each instance associated with a respective
rule set 435 and, preferably, a respective dictionary 437,
or other lexical analysis tool. A respective instance
436' of the text analyzer 436 is provided for each root or
branch node 311, 312 of the hierarchical rule structure
10 310 being implemented (and therefore for each rule set).

The respective rule set for each instance 436' of the text
analyzer 436 therefore corresponds with the rule set for
the respective root or branch node 311, 312.

15 The classifier 434 conveniently, but not necessarily,
takes the form of a CORBA server, preferably written in
Java, or similar programming language. The text analyzer
36 is conveniently, but not necessarily, available to the
classifier 434 as part of a DLL (Dynamic Link Library), or
20 a UNIX SO (shared object) as appropriate, and arranged to
allow multiple instantiation.

The classifier 434 is arranged to retrieve electronic
messages 12 from a database 432. Conveniently,
25 communication between the classifier 434 and the database
432 is effected by a database server 440 in conventional
manner. The database server 440 also provides the
classifier 434 with access to a file system 442 which
conveniently stores the rule sets, the hierarchical rule
30 structure to be implemented and, if applicable, one or
more dictionaries (or other lexical analysis tool).

The operation of the classifier 434 is now described with reference to Figure 5. At module 501, the classifier 434 retrieves from the database 432 details of nodes in the hierarchical structure to be implemented, together with 5 the respective rule sets for each node and, if lexical analysis is desired, a respective dictionary, or other lexical analysis tool. At module 502, the classifier 434 retrieves from the database 432 details of the inter-dependencies of the nodes in the hierarchical structure. 10 All of the information retrieved at modules 501 and 502 is conveniently stored in memory by the classifier 434.

At module 503, the classifier 434 instantiates a respective instance 436' of the text analyzer 436 for each 15 root and branch node 311, 312 and associates with each instance 436' the respective corresponding rule set and, if applicable, dictionary.

At module 504, the classifier 434 retrieves from the 20 database 432 a message 12 to be classified. At module 505, the retrieved message is provided to the instance 436' of the text analyzer 436 which corresponds to the root node 311 of the hierarchical rule structure. The contents of message 12 are evaluated by said instance 436' 25 of the text analyzer 436 and a result is returned to the classifier 434.

At module 506, the classifier 434 causes the text 30 analyzer, in conjunction with its rule engine, to analyze the message at hand according to the rule set associated with the current node (initially the root node). There are two possible outcomes from the analysis. The first is that the text analyzer 436 produces a set of one or more

results which have been generated by application of the rule set to the message. In the present example, the or each result may indicate that the message either belongs to a particular category, or that the message should next 5 be analyzed by the rule set of a further branch node 312. It will be appreciated that, in alternative embodiments, a result may have other indications depending on the system within which the text analyzer is intended to operate. Since the text analyzer 436 may produce multiple results, 10 it follows that the text analyzer 436 may determine that the message belongs to one or more categories and/or should be analyzed in accordance with one or more further branch nodes 312. With reference to Figure 3, this effectively means that a message may be routed through the 15 hierarchical structure 310 along more than one path. Each path corresponds with a respective sequence of rule sets under which the message is evaluated. The second possible outcome of the analysis is that the text analyzer 436 does not produce any results i.e. the result set is a null set. 20 In this case the classifier 434 implements a default option as described below.

Module 507 tests whether or not the analysis at module 506 has produced any results. If one or more results have 25 been produced then the classifier 434 proceeds to module 508 where it tests whether or not there is a result to process (which there will always be when the classifier 434 reaches module 508 from module 507). When module 508 determines that there is a result to be processed, the 30 classifier 434 proceeds to module 509 where it tests whether or not the result indicates that the message belongs to a particular category i.e. whether or not a leaf node 316 of the hierarchical structure 310 being

implemented has been reached. If so, then the classifier 434 proceeds to module 510 where it stores a record of the category indicated by the result. The category may be stored in any conventional manner - conveniently in a file 5 for association with the message being analyzed. The classifier 434 then returns to module 508 to determine whether or not there are any further results to be processed.

10 Should module 509 determine that the result is not indicative of a message category, the classifier 434 proceeds to module 511 where it tests whether or not the result indicated that the message should be analyzed in accordance with a further branch node 312. If so, then 15 the classifier 434 routes the message to the branch node 312 indicated by the result, i.e. it passes the message to the text analyzer instance 436' associated with the indicated branch node 312, and returns to module 506 so that modules 506 to 511 are repeated, only this time by a 20 different text analyzer instance using a different rule set.

Also from module 511, the classifier 434 returns to module 508 to determine whether or not there are any further 25 results to be processed.

Thus, the operation of the classifier 434 is recursive in nature and may, for example, be performing modules 508 to 511 using the text analyzer instance and rule set 30 associated with a first node while also performing modules 506 to 511 using the text analyzer and rule set associated with a second node, and so on, all on the same message. The dashed lines in Figure 5 are used to indicate where

recursion may occur. The recursive nature of the classifier 434 arises because of the possibility of navigation through the hierarchical structure by more than one route.

5

If, at module 507, it is found that no results are produced by the text analyzer, then advantageously a default procedure is followed. In the preferred embodiment, the default procedure is to assign the message to a pre-determined default node or category. In the case of a default node being desired, the classifier 434, upon return to module 511 from module 507, proceeds to module 506 and so routing the message to the text analyzer associated with the default node. If a default category is desired, then at module 511 the classifier 434 proceeds to module 510 and stores the default category.

Similarly, should the classifier 434 reach module 511 from module 509 and determine that the correct result does not 20 match any valid nodes, then the default procedure is followed.

The classifier 434 terminates its processing of the message when, at module 508, it determines that there are 25 not further results to process.

The classifier 434 returns classified messages to the database 432, together with the determined category information and, preferably, information on how the 30 message should be routed in view of its classification, for distribution by the distributor 40.

The classifier 434 need not necessarily launch multiple instances of the text analyzer 436 simultaneously (module 503). In an alternative embodiment, the classifier 434 may be arranged to instantiate an instance 436' of the 5 text analyzer 434 corresponding to a particular node only when it is determined by the rules evaluation under a preceding node that evaluation under said particular node is required.

10 The advantages of the invention can readily be appreciated by reference to the example of Figure 4. In this example it is assumed that a banking organisation has two sub-organizations, one in the USA and one in Germany, and can therefore receive electronic messages 12 in either the 15 English language or the German language. It is further assumed that each sub-organization has two divisions namely, business banking and retail banking. Each division has a list of categories in which to place incoming electronic messages 12, some of which they may 20 have in common.

The text analyzer instance 436' which corresponds to the root node 311 is designated LANGUAGE and is associated with its own set of rules 435 and a dictionary. The 25 purpose of the root node 311 in this example is to determine whether the message 12 to be classified is in English or in German. To do this, the rule set 435 associated with the LANGUAGE instance 436' is arranged to search for words (sometimes referred to as stop-words) in the message 12 which can readily be used to identify a 30 language. In this case, "der", "das", "the" or "it" may be suitable stop-words to search for.

Two further instances 436' of the text analyzer 436 are designated ENGLISH and GERMAN respectively. The ENGLISH instance corresponds to a branch node 312 under which the message 12 is to be evaluated if the LANGUAGE instance 5 determines that the message 12 is in English. The GERMAN instance corresponds to a branch node 312 under which the message 12 is to be evaluated if the LANGUAGE instance determines that the message 12 is in German. Each of the ENGLISH and GERMAN instances are associated with a 10 respective rule set 435 and with an English dictionary 437 and a German dictionary 437 respectively.

The purpose of the respective rule sets 435 associated with the ENGLISH and GERMAN instances is the same, namely 15 to establish whether a message 12 relates to the retail banking division or the business banking division of the respective sub-organization. However, the individual rules within each rule set 435 are likely to be different as they may depend on, for example, the language, culture 20 and local practice of the respective sub-organizations i.e. a rule set generated from a batch of sample messages received by a German sub-organization and submitted by German customers will not be the same as a rule set generated from a batch of sample messages received by a 25 U.S. sub-organization and submitted by U.S. customers.

At this point, an advantage of a hierarchical rule structure in which more than one rule set exists becomes readily apparent - the U.S. organization can create, 30 maintain and amend its own rule set 435 for distinguishing between retail-related and business-related messages 12 without compromising the equivalent rule set 435 held by the German sub-organization, and vice versa.

A further advantage is also apparent in that the benefits of lexical analysis, or word-stemming, are not lost to either the U.S. or German sub-organization since all
5 messages 12 are analyzed by a text analyzer 436 which is associated with a dictionary in the same language as the message 12 itself.

A further four instances of text analyzer 436 are
10 designated ENGLISH RETAIL, ENGLISH BUSINESS, GERMAN RETAIL and GERMAN BUSINESS respectively, each being associated with a respective set of rules 435 and an English or German dictionary 437, as appropriate. The ENGLISH RETAIL instance corresponds to a branch node 312 under which the
15 message 12 is to be evaluated if the ENGLISH instance determines that the message 12 relates to the retail banking division. The ENGLISH BUSINESS instance corresponds to a branch node 312 under which the message 12 is to be evaluated if the ENGLISH instance determines
20 that the message 12 relates to the business banking division. The GERMAN RETAIL instance corresponds to a branch node 312 under which the message 12 is to be evaluated if the GERMAN instance determines that the message 12 relates to the retail banking division. The
25 GERMAN BUSINESS instance corresponds to a branch node 312 under which the message 12 is to be evaluated if the GERMAN instance determines that the message 12 relates to the business banking division.
30 In this example, it is assumed that these latter four instances correspond to the final branch nodes in the hierarchical rule structure being implemented and that their purpose, therefore, is to determine into which

category, or categories, to classify a message 12. Example possible categories include "balance enquiry", "request for fund transfer", "interest rate enquiry" and so on. The respective list of categories for the U.S.

5 retail banking division may, or may not, be identical to those of the German retail banking division, but there are likely to be at least a few categories in common. Similar comments apply in relation to the respective business banking divisions. The advantage of the hierarchical rule

10 structure can again be appreciated since each division of each sub-organization can administer its own rule set 435 in isolation from the other rule sets and therefore without interfering with the other rule sets. Such interference would be unavoidable if a conventional flat

15 rule structure was employed since each sub-organization/division would inevitably generate conflicting rules for identifying the category, or categories, they have in common.

20 It will further be apparent that the hierarchical rule structure of the invention significantly reduces the number of rule evaluations which any given message 12 must undergo. For example, in Figure 4, a message 12 which relates to the retail banking division of the U.S. sub-

25 organization does not have to be subjected to any of the rules which are particular to the German sub-organization or to the U.S. business banking division. This represents a significant saving in the time taken to process messages 12 and reduces the processing power required by the

30 organization.

The invention is described above in the context of a simple banking organization. It will be appreciated,

however, that the invention is equally suitable for application other fields.

Also, the invention is described herein in the context of processing electronic messages (such as e-Mails) asynchronously but it will be appreciated that the invention is also suitable for processing electronic messages synchronously or in real-time.

10 The invention is not limited to the embodiments described herein which may be modified or varied without departing from the scope of the invention.

CLAIMS:

1. An electronic message processing system arranged to receive electronic messages, the system comprising:

5

means for storing a plurality of rules;

a text analyzer;

10 a rule engine associated with the text analyzer and with the rule storage means,

the text analyzer and the rule engine being co-operable to apply one or more rules to the content of a received 15 electronic message and to generate one or more results based on the application of the or each rule;

a classification module co-operable with the text analyzer and the rule engine and arranged to classify the 20 electronic message into one or more message categories based on said one or more results,

wherein the rules are arranged into a plurality of rule sets, the classification module being arranged to cause 25 the text analyzer in association with rule engine to apply one or more rule sets to the message content in accordance with a hierarchical structure whereby the or each result generated by application of one rule set to the message content determines the next rule set, or rule sets, to be applied.

30 2. An electronic message processing system as claimed in Claim 1, wherein the text analyzer and the rule engine are

arranged to generate a respective result set for the or each rule set applied to the message content, the classification module being arranged to determine respectively from the or each result of the or each result 5 set whether to classify the message in a category or to cause a further rule set to be applied to the message content.

3. An electronic message processing system as claimed in 10 claim 1, wherein the a text analyzer includes the rule engine.

4. An electronic message processing system as claimed in 15 claim 3, wherein the classification module is arranged to instantiate a respective instance of the text analyzer for each rule set, each text analyzer instance being arranged to apply its respective rule set to the message content.

5. An electronic message processing system as claimed in 20 claim 4, wherein each text analyzer instance is associated with a respective lexical analysis tool.

6. An electronic message processing system as claimed in 25 claim 5, wherein the lexical analysis tool includes a dictionary.

7. An electronic message processing system as claimed in 30 claim 1, wherein the rule storage means comprises a plurality of rule files, each rule file containing a respective rule set.

8. A classification module for use in an electronic message processing system, the system comprising means for

storing a plurality of rules; a text analyzer; a rule engine associated with the text analyzer and with the rule storage means, the text analyzer and the rule engine being co-operable to apply one or more rules to the content of
5 an electronic message received by the system and to generate one or more results based on the application of the or each rule,

10 the classification module being arranged for co-operation with the text analyzer and the rule engine and further arranged to classify the electronic message into one or more message categories based on said one or more results,

15 wherein the rules are arranged into a plurality of rule sets, the classification module being arranged to cause the text analyzer in association with the rule engine to apply one or more rule sets to the message content in accordance with a hierarchy whereby the or each result generated by the of one rule set the message content
20 determines the next rule set, or rule sets, to be applied.

9. In an electronic message processing system arranged to receive electronic messages, the system comprising
25 means for storing a plurality of rules; a text analyzer; a rule engine associated the text analyzer and with the rule storage means, the text analyzer and the rule engine being co-operable to apply one or more rules to the content of a received electronic message and to generate one or more
30 results based on the application of the or each rule; and a classification module co-operable with text analyzer and the rule engine and arranged to classify the electronic message into one or more message categories based on said

one or more results, a method of classifying an electronic message comprising:

arranging the rules into a plurality of rule sets;

5

causing the text analyzer, in association with the rule engine, to apply one or more rule sets to the message content in accordance with a hierarchy; and

10 determining the next rule set, or rule sets, to be applied depending on the or each result generated by application of the preceding rule set to the message content.

15 10. A method of classifying an electronic message as claimed in claim 9, further including:

instantiating a respective instance of the text analyzer for each rule set; and

20 20. arranging each text analyzer instance to apply its respective rule set to the message content.

11. An electronic message processing system as claimed in claim 1, wherein the electronic messages to be processed 25 include unstructured text-based messages.

12. An electronic mail (e-mail) processing system comprising an electronic message processing system as claimed in claim 1.

30

13. An SMS message processing system comprising an electronic message processing system as claimed in claim 1.

ABSTRACT

Electronic Message Processing

5 An electronic message processing system includes a classification module for classifying electronic messages, such as e-mails and SMS messages, into one or more categories according to their content. In a preferred 10 embodiment, a text analyzer is used to apply a sequence of one or more rule sets to the content of each message. The rule sets to be applied and their order in the sequence is determined by a hierarchical structure. The use of the hierarchical structure reduces the number of rules which 15 must be applied to a given message. Arranging the rules in sets facilitates administration and maintenance of the rules. A further aspect of the invention involves the instantiation of multiple instances of the text analyzer, one instance for each rule set. This allows an 20 appropriate lexical analysis tool, such as a dictionary in an appropriate language, to be used in conjunction with each rule set.

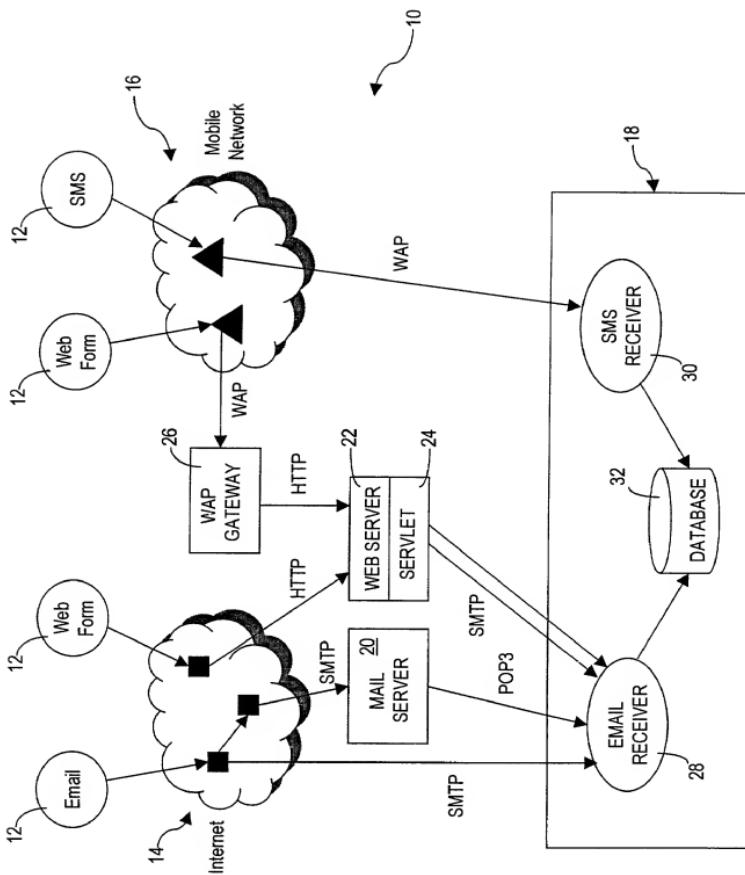


Fig.1

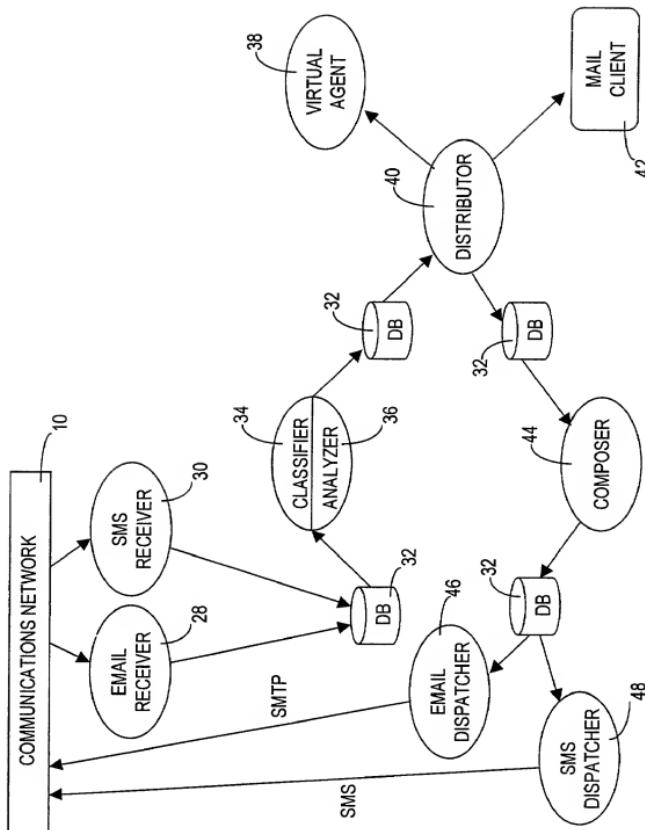


Fig. 2

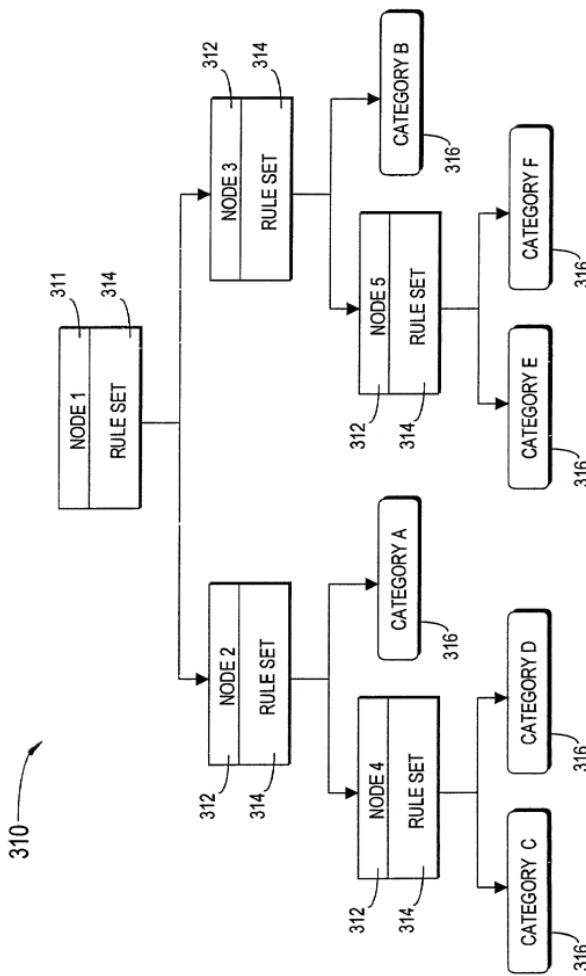


Fig. 3

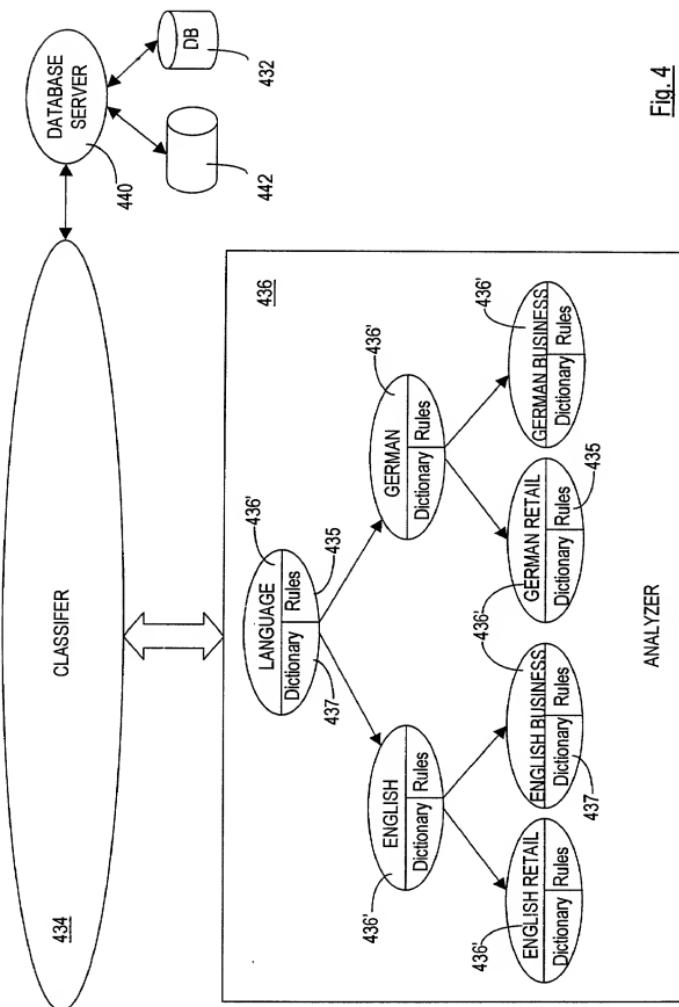


Fig. 4

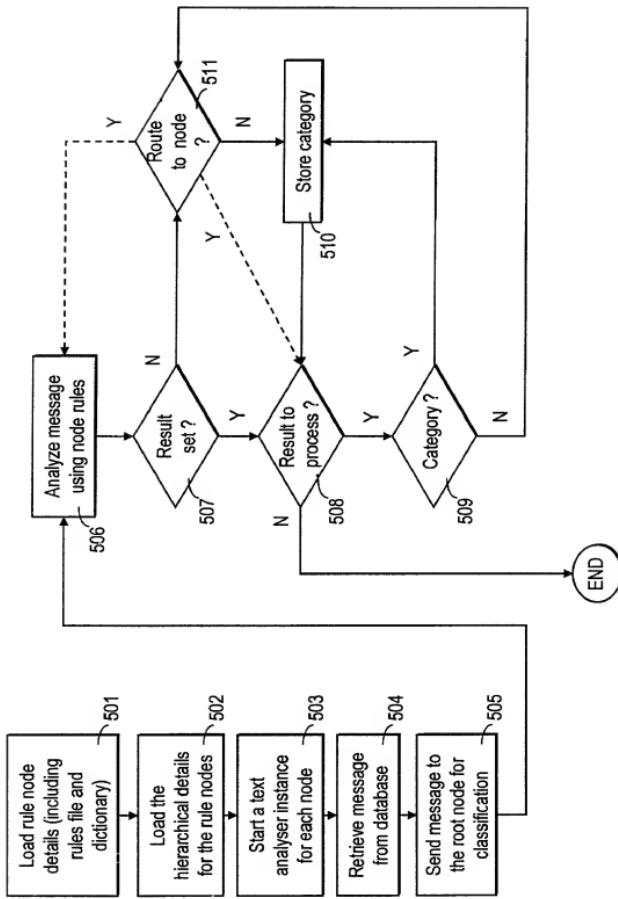


Fig. 5

Declaration and Power of Attorney For Patent Application

English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

ELECTRONIC MESSAGE PROCESSING

the specification of which is attached hereto unless the following box is checked:

was filed on _____ as

United States Application Number or PCT International Application Number _____

and was amended on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR § 1.56.

I hereby claim foreign priority benefits under 35 U.S.C. §119(a)-(d) or § 365(b) of any foreign application(s) for patent or inventor's certificate, or § 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below by checking the box, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)

Priority Not Claimed

I hereby claim the benefit under 35 U.S.C. § 119(e) of any United States provisional application(s) listed below.

I hereby claim the benefit under 35 U.S.C. § 120 of any United States application(s), or 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. § 112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR § 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application:

(Application Number) (Filing Date) (Status - patented, pending, abandoned)

(Application Number) (Filing Date) (Status - patented, pending, abandoned)

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith:

Paul F. Prestia	Reg. No. 23,031	Lawrence E. Ashery	Reg. No. 34,515	Jacques L. Etkowicz	Reg. No. 41,738
Allan Ratner	Reg. No. 19,717	Christopher R. Lewis	Reg. No. 36,201	Jack J. Jankovitz	Reg. No. 42,690
Andrew L. Ney	Reg. No. 20,300	Robert L. Andersen	Reg. No. 25,771	Jonathan H. Spadt	Reg. No. 45,122
Kenneth N. Nigon	Reg. No. 31,549	Joshua L. Cohen	Reg. No. 38,040	Christopher I. Halliday	Reg. No. 42,621
Kevin R. Casey	Reg. No. 32,117	Daniel N. Calder	Reg. No. 27,424	Scott A. McKeown	Reg. No. 42,866
Benjamin E. Leace	Reg. No. 33,412	Louis W. Beardell, Jr.	Reg. No. 40,506		
James C. Simmons	Reg. No. 24,842				

Address all correspondence to: Kenneth N. Nigon

Ratner & Prestia, Suite 301, One Westlakes, Berwyn, P.O. Box 980, Valley Forge, PA 19482-0980

Address all telephone calls to: Kenneth N. Nigon at (610) 407-0700.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor (given name, family name) Patrick McErlean

Inventor's signature _____ Date _____

Residence 14 Ballymacpeake Road, Portglenone, County Antrim, Northern Ireland, United Kingdom

Citizenship United Kingdom

Post Office Address 14 Ballymacpeake Road, Portglenone, County Antrim, Northern Ireland, United Kingdom

Additional inventors are being named on separately numbered sheets attached hereto.